Discrete Mathematics 93 (1991) 55–61
North-Holland

On a tree-cutting problem of P. Ash

I. Krasikov
School of Mathematical Sciences, Tel-Aviv University, Israel

Received 12 January 1988
Revised 2 November 1989

Abstract

Krasikov, I., On a tree-cutting problem of P. Ash, Discrete Mathematics 93 (1991) 55–61.

It is shown that in every tree T with N vertices, there are k vertices such that the connected components obtained by deleting those k vertices can be partitioned into two classes C_1^k and C_2^k with

$$\delta_k(T) = ||C_1^k| - |C_2^k|| \leq \max\left(1, 3^{-k}\left(N - \frac{3^k - 1}{2}\right)\right).$$

Moreover, for each k there is an infinite family of trees, namely, ternary trees, with the optimal value of $\delta_k(T)$ equal to

$$3^{-k}\left(N - \frac{3^k - 1}{2}\right) - 2(k - 1).$$

We also consider the corresponding question for the edge deletion.

1. Introduction

In the Ninth British Combinatorial Conference, P. Ash asked the following question [1].

> One wishes to split a tree into two parts with an approximately equal number of vertices in each part. First, he deletes k vertices and puts each connected component of the obtained forest into one of two classes. How well can he do it?

More formally, let T be a tree and let $D_k = \{v_1, v_2, \ldots, v_k\}$ be a subset of its vertices. Furthermore, let the connected components of $T - D_k$ be split into two classes C_1^k and C_2^k with $|C_1^k|$ and $|C_1^k|$ vertices respectively such that $\delta_k(T) = ||C_1^k| - |C_2^k||$ is as small as possible. Put $F_k(T) = \min_{D_k \subset T} \delta_k(T)$. What is $R_k(N) = \max_{|T|=N} F_k(T)$?

It is easy to see that $R_1(N)=[(N-1)/3]$, where $[x]$ means the largest integer not exceeding x. But already for $k=2$ the situation is not so simple. Bermond proved $R_2(N)<N/7$, and it was conjectured that $R_2(N)\leq N/9$ [2].

Here we will establish

$$3^{-k}\left(N-\frac{3^k-1}{2}\right)-2(k-1)\leq R_k(N)\leq \max\left(1,3^{-k}\left(N-\frac{3^k-1}{2}\right)\right).$$

Choosing D_k to be a subset of edges, one can also study an edge deletion variant of the above problem. Here it is convenient to consider trees with the maximum degree not exceeding Δ, to avoid a trivial extremal example given by a star $K_{1,N-1}$. So, the corresponding error term, $r_k(N,\Delta)$, turns out to be a function of two variables.

Ash and Daykin [2]. proved that $r_k(N,3)\leq 1$, whenever $k>\log_{3/2}N$. Here we will show

$$r_k(N,\Delta)\leq\left[\left(\frac{\Delta-2}{\Delta}\right)^k(N-1)\right]+1.$$

2. Results

First, we establish an upper bound for $R_k(N)$.

Theorem 1.

$$(1)\qquad R_k(N)\leq\max(1,t_k),\quad \textit{where } t_0=N,\ t_{i+1}=\left[\frac{t_i-1}{3}\right].$$

In particular,

$$(2)\qquad R_k(N)\leq\max\left(1,3^{-k}\left(N-\frac{3^k-1}{2}\right)\right).$$

Proof. The proof is by induction on k. By definition, $R_0(N)=t_0=N$. Note that the numbers $\delta_k=\delta_k(T)$ are alternately odd and even. Moreover, if $\delta_k=0$ or 1 for some k then $\delta_{k+1}=1$ or 0 respectively. Assume now that D_{k-1} and the corresponding partition $P_{k-1}=C_1^{k-1}\cup C_2^{k-1}$ satisfying (1) have already been obtained.

Suppose first that $\delta_{k-1}\leq 4$ then the required partition P_k can be obtained as follows:

if $\delta_{k-1}=0$ one deletes some vertex,

if δ_{k-1} is 1 or 2 one deletes a leaf from the large class.

if δ_{k-1} is 3 or 4 one deletes a vertex adjacent to a leaf x in the large class and send the now isolated vertex x in the small class.

Now suppose that $\delta_{k-1} \geqslant 5$. Observe that for any connected component Y belonging to the bigger class, say C_2^{k-1}, we may assume $|Y| \geqslant \delta_{k-1}$ since otherwise a partition can be improved by moving Y into C_1^{k-1}.

In order to obtain P_k we delete one vertex v from Y and divide the obtained components into two parts Y' and Y''. After moving Y' into C_1^{k-1} we have

$$\delta_k = |\delta_{k-1} - 1 - 2|\,Y'||.$$

Now, we show that v may be chosen so that for δ_k (1) holds.

Select both a vertex v of Y and a partition $A \cup B$ of $Y - v$ giving A as small as possible and

$$(3) \qquad |A| \geqslant \frac{\delta_{k-1} - 1}{2}.$$

Let (v, A) be such a selection. Such selections exist since $|Y| > \delta_{k-1}$.

If $|A| = (\delta_{k-1} - 1)/2$ holds then $\delta_k = 0$. If $|A| = \delta_{k-1}/2$ holds then $\delta_k = 1$. Thus one may assume $|A| \geqslant (\delta_{k-1} + 1)/2$, and then A is not connected, since otherwise one could move v to an adjacent position in A, thus decreasing the size of A by one.

Let $a_1, a_2, \ldots, a_l$ be the connected components of A and $|a_1| \leqslant |a_2| \leqslant \cdots \leqslant |a_l|$. Then

$$\sum_{i=2}^{l} |a_i| < \frac{\delta_{k-1} - 1}{2} < \sum_{i=1}^{l} |a_i|.$$

Indeed, the first inequality holds, since otherwise one could replace A by $A \setminus a_1$. Now if

$$\sum_{i=2}^{l} |a_i| \geqslant \frac{\delta_{k-1} - 1}{3}$$

then

$$\delta_{k-1} - 1 - 2\sum_{i=2}^{l} |a_i| \leqslant \frac{\delta_{k-1} - 1}{3} \leqslant \frac{t_{k-1} - 1}{3}.$$

Otherwise,

$$\sum_{i=1}^{l} |a_i| \leqslant \frac{2(\delta_{k-1} - 1)}{3}$$

and we have

$$2\sum_{i=1}^{l} |a_i| + 1 - \delta_{k-1} \leqslant \frac{\delta_{k-1} - 1}{3} \leqslant \frac{t_{k-1} - 1}{3}.$$

Hence the required partition is obtained by moving either $\bigcup_{i=2}^{l} a_i$ or $\bigcup_{i=1}^{l} a_i$ into C_1^{k-1} depending on whether the first or the second inequality holds. $\square$

Now we will obtain a lower bound on $R_k(N)$ showing that (2) is almost sharp. First, we need a few more definitions.

The ternary tree T_L with L levels is defined as follows. The first level of T_L is just a single vertex w. Each vertex v from the ith level, $i < L$, is adjacent to three vertices of the $(i+1)$th level and to the only vertex of the $(i-1)$th level, provided $v \neq w$.

We denote by v_i^+, $i = 0, 1, 2$, three succeeding vertices and by v^- the unique preceding vertex to v respectively. The tree consisting of v and all vertices succeeding v will be denoted by $T(v)$. Thus, if v is on the ith level, then $|T(v)| = (3^{L-i+1} - 1)/2$.

Theorem 2. *Let T_L be a ternary tree with L levels and $N = (3^L - 1)/2$ vertices. Then for $1 \leqslant k \leqslant L/2$,*

$$F_k(T_L) \leqslant \left|3^{-k}\left(N - \frac{3^k - 1}{2}\right) - 2(k-1)\right| = \left|\frac{3^{L-k} - 1}{2} - 2(k-1)\right|.$$

Proof. Let us construct a required partition of T_L. For, let s_j be a string of length $j+1$, consisting of j ones followed by a zero. Let s be a string consisting of $k-1$ ones. To each edge (v, v_i^+) of T_L we assign the number i. Each vertex $v \in T_L$, $v \neq w$, will be identified with string $s(v)$, where the entries of $s(v)$ are the labels given to the edges on the oriented path $w - v$.

Select the vertices corresponding to the strings $s \cup (\bigcup_{j=0}^{k-2} s_j)$ to be the vertices of D_k. Observe that $T(v) \cap T(u) = \emptyset$ for any two different $v, u \in D_k$. Set

$$C_2^k = \left(\bigcup_{v \in D_k} (T(v) \backslash v)\right) \backslash T(u_0^+),$$

where u is the vertex with string s. Note that if $s(v)$ has l entries then $|T(v)| = (3^{L-l} - 1)/2$. Now a simple calculation shows that this is the required partition. $\square$

Theorem 3. *If $2k \leqslant L$ then*

$$(4) \qquad F_k(T_L) = \frac{3^{L-k} - 1}{2} - 2(k-1).$$

Proof. For the sake of brevity we will write here f_k instead of $F_k(T_L)$. In virtue of Theorem 2 it is left to show that

$$(5) \qquad F_k \geqslant \frac{3^{L-k} - 1}{2} - 2(k-1).$$

The proof is by induction on k. Obviously, (4) holds for $k = 1$. So, we may assume $k \geqslant 2$. Moreover, in virtue of the assumption $L \geqslant 2k$, we also assume that

$$F_{k-1} \geqslant \min_{k \geqslant 2}\left(\frac{3^{k+1} - 1}{2} - 2(k-2)\right) = 13.$$

Let u be a vertex of the highest level h in D_k, and let M_h be the tree $T(u_i^+)$, $i \in \{0, 1, 2\}$. We will not distinguish between different M_h corresponding to a particular choice of u and i. Notice also that any three M_h, belonging to the same class, may be regarded as having a common predecessor u in T_L. This will be widely used throughout the proof.

Let $x \in D_k$ and let either $x^- \in C_i^k$ or $x^- \in D_k$. Denote by $P_{k-1}^i(x)$ a partition obtained from the initial partition P_k by returning x and moving $T(x)$ into C_i^k.

To establish (5), we first prove

$$(6) \qquad |M_h| \geqslant \frac{3^{L-k}-1}{2},$$

which implies that $h \leqslant k$.

Assume the contrary. Without loss of generality, suppose $|C_1^k| \leqslant |C_2^k|$. Denote by μ the number of M_h occurring in C_1^k. We shall consider three following cases.

Case (i): $\mu \geqslant 3$.

If $u^- \in D_k$ or $u^- \in C_1^k$ then (5) holds. Indeed, we should have $F_k = 0$, $F_{k-1} = 1$, since otherwise, by returning u we get $F_{k-1} < F_k$. This, obviously, contradicts $F_{k-1} \geqslant 13$. If $u^- \in C_2^k$ then we could improve our partition replacing u by u^- in D_k, unless $F_k \leqslant 1$. If $F_k \leqslant 1$ then we consider $P_{k-1}^2(u)$ and the corresponding value of δ_{k-1}. We have $F_{k-1} \leqslant \delta_{k-1} = F_k + 6\,|M_h| + 1$. Hence, by the induction hypothesis,

$$\frac{3^{L-k+1}-1}{2} - 2(k-2) \leqslant 3^{L-k}.$$

Thus, $3^{L-k} \leqslant 4k - 9$. On the other hand,

$$F_k \leqslant \frac{3^{L-k}-1}{2} - 2(k-1).$$

Hence,

$$3^{L-k} \geqslant 4k - 3 + 2F_k > 4k - 9.$$

This is impossible and so, (6) holds.

Case (ii): $1 \leqslant \mu \leqslant 2$.

If $u^- \in D_k$ or $u^- \in C_2^k$ then we consider $P_{k-1}^2(u)$. We get $F_{k-1} \leqslant F_k + 4\,|M_h| + 1$. But $F_k \leqslant |M_h|$, so, $F_{k-1} \leqslant 5\,|M_h| + 1$. This leads to a contradiction, as in the previous case.

The case $u^- \in C_1^k$ can be considered in a similar way. We omit the details.

Case (iii): $\mu = 0$.

If $u^- \in C_2^k$, then $F_{k-1} \leqslant 1$ and we are ready. Otherwise, $P_{k-1}^1(u)$ yields $F_{k-1} \leqslant 6\,|M_h| + 1 - F_k$. Now, as above, we see that (6) must hold.

Further, we claim that an optimal partition, corresponding to F_k, satisfies the following conditions:

(a) $h = k$;

(b) $u^- \in C_1^k$;

(c) There is only one $M_k \subset C_1^k$:

(d) There are at least two vertices, u and, say, u_1, from the kth level in D_k.

Let us prove (a)–(d). Denote by d_l the difference between the number of vertices from the lth level in C_2^k and C_1^k respectively. Observe that the number of vertices on each level of T_L is odd. Since u is on the $h \leqslant k$ level then

$$|S_h| = \left|\sum_{l=h+1}^{L} d_l\right| \geqslant |T_{L-h}| = |M_h|.$$

Moreover, $|S_h| = S_h = \sum_{l=k+1}^{L} d_l = |M_k|$, and, so, (a) holds. Indeed, otherwise $|S_h| \geqslant 2|M_k|$ by (6). But $\sum_{l=1}^{h} |d_l| \leqslant |T_h| - k$, hence,

$$\frac{3^{L-k}-1}{2} - 2(k-1) \geqslant F_k \geqslant 2\,|M_k| - |T_h| + k.$$

One can check that this contradicts the assumption $L \geqslant 2k$ and (6).

To prove (b) assume $u^- \in C_2^k$. Then $P_{k-1}^2(u)$ gives

$$F_{k-1} \leqslant F_k + 2\,|M_k| + 1 \leqslant \frac{3^{L-k+1}-1}{2} - 2(k-1),$$

in contradiction with the induction hypothesis. Thus, (b) also holds.

Let us demonstrate (c). We may regard any two M_k occurring in C_1^k as having a common predecessor u in T_L. Again, we consider $P_{k-1}^1(u)$. Then $F_{k-1} \leqslant 2\,|M_k| + 1 - F_k$ which is impossible.

Finally, if (d) is false, then C_1^k contains 7 (mod 9) subtrees T_{L-k} on the $(k+1)$th level by (c). This implies $S_k \geqslant 2\,|M_k|$, which has already been shown to be impossible.

Now we are ready to complete the proof. Since $(u_1)^-$ has to be also in C_1^k, we may assume that $(u_1)^- = u^-$. Observe that either $T(u)\backslash u$ or $T(u_1)\backslash u_1$ is in C_2^k. We replace u and u_1 by u^- in D_k and move $T(u)$ and $T(u_1)$ into C_2^k. Then $F_{k-1} \leqslant F_k + 2\,|M_k| + 3$, and (5) follows. This completes the proof. □

Remark. Probably the condition $L \geqslant 2k$ is superfluous and (4) holds whenever its right-hand side is at least 1.

Conjecture.

$$R_k(N) \leqslant \max\left(1, 3^{-k}\left(N - \frac{3^k-1}{2}\right) - 2(k-1)\right).$$

For the case of edge deletion an upper bound on the corresponding error term $r_k(N, \Delta)$ is given by the following theorem.

Theorem 4.

$$(7) \qquad r_{k+1}(N, \Delta) \leqslant \left\lceil \frac{(\Delta-2)r_k(N, \Delta) + 2}{\Delta} \right\rceil.$$

In particular,

$$(8)\qquad r_k(N, \Delta) \leqslant \left[\left(\frac{\Delta-2}{\Delta}\right)^k (N-1)\right] + 1.$$

Proof. We begin with the following observation. Let H be a tree with the maximum degree Δ. Fix a number c, $0 < c \leqslant |H|$. Then there is an edge $e \in H$, such that

$$\frac{c-1}{\Delta-1} \leqslant |S| \leqslant c,$$

for some component $S \subset H \backslash e$.

Indeed, select a vertex v and a branch C at v such that all branches at v (but for one perhaps) have at most c vertices (v not included); the branch is as large as possible. Cutting this branch near v gives a component with at most c vertices. Since the cutting of the last edge incident to v gives a component containing v with more than c vertices (otherwise the selection would not have maximal size), our branch has at least $(c-1)/(\Delta-1)$ vertices.

Now consider an optimal splitting C_k^1, C_k^2, and the corresponding value of $\delta_k(T) = ||C_k^1| - |C_k^2||$ of a tree T. Let A be a minimal component of a bigger class, say C_k^2. Then, of course, $\delta_k(T) \leqslant |A|$.

Choose

$$c = \frac{(\Delta-1)\delta_k(T)+1}{\Delta}.$$

Delete an edge from A so that the size of one of the obtained components, A_1, is in the interval

$$\left[\frac{c-1}{\Delta-1}, c\right].$$

Moving A_1 into C_k^1 yields a partition satisfying

$$\delta_{k+1}(T) \leqslant \frac{(\Delta-2)\delta_k(T)+2}{\Delta}.$$

Hence (7) holds. Solving this recurrence yields (8). □

References

[1] P. Ash, Ninth. British Combinatorial Conference. Southampton, 1983.

[2] P. Ash and Bill Jackson, private communication.